

On quadrature methods for refractory point process likelihoods

Gonzalo Mena and Liam Paninski

Statistics Department and Grossman Center for the Statistics of Mind
Columbia University

July 2, 2014

Abstract

Parametric models of the conditional intensity of a point process (e.g., generalized linear models) are popular in statistical neuroscience, as they allow us to characterize the variability in neural responses in terms of stimuli and spiking history. Parameter estimation in these models relies heavily on accurate evaluations of the log-likelihood and its derivatives. Classical approaches use a discretized time version of the spiking process, and recent work has exploited the existence of a refractory period (during which the conditional intensity is zero following a spike) to obtain more accurate estimates of the likelihood. In this brief note we demonstrate that this method can be improved significantly by applying classical quadrature methods directly to the resulting continuous-time integral.

1 Introduction

Point processes are a powerful theoretical tool to characterize the occurrence of random events in time, and are widely used in statistical neuroscience (Truccolo, 2010) to represent neural spiking in terms of intrinsic factors, including refractoriness and the activity of other neurons, and extrinsic factors, such as stimulus-driven effects. One way to characterize such processes is through the conditional intensity function (CIF). Intuitively, the CIF can be interpreted as an instantaneous firing rate given the past history. This history, denoted \mathcal{H}_t , is meant to represent changes in spiking probabilities due to previous activity, e.g., refractoriness. The log-likelihood of observing the spikes at times $t_1, \dots, t_{N(T)}$ in the interval $[0, T]$ is given by (Brillinger, 1988; Daley & Vere-Jones, 2007)

$$l(t_1, \dots, t_{N(T)}) = \sum_{i=1}^{N(T)} \log(\lambda(t_i | \mathcal{H}_{t_i})) - \int_0^T \lambda(t | \mathcal{H}_t) dt, \quad (1)$$

with $\lambda(t|\mathcal{H}_t)$ denoting the CIF at time t .

Usually, an explicit functional model of the CIF is assumed. The most popular approach involves generalized linear models (GLM) (Brillinger, 1988; Paninski, 2004; Truccolo, Eden, Fellows, Donoghue & Brown, 2005), in which the CIF is a function of a linear transformation of the stimulus and other known covariate terms. This leads to a parametric expression for the likelihood, in which maximum likelihood (ML) or maximum a posteriori (MAP) estimates are computed using standard optimization methods that evaluate the log-likelihood (and derivatives) at each iteration. The main problem is that only approximations of the full continuous-time log-likelihood can be computed, as there may not be a closed form expression for the cumulative CIF, $\int_0^T \lambda(t|\mathcal{H}_t) dt$.

Previous approaches are based on discrete-time versions of the spiking process, which in turn lead to discretizations of the log-likelihood that converge to the continuous version when the bin size goes to zero. For point processes with refractory effects, in which the CIF jumps discontinuously to zero following a spike, (Citi, Ba, Brown & Barbieri, 2014) recently developed an improved such discretization. Here we show that an alternative approach, in which we apply standard quadrature methods directly to the original continuous time integral, leads to significant further improvements beyond those offered by the approach of (Citi et al., 2014), with minimal additional computational cost¹.

2 A direct quadrature method for approximating the continuous-time log-likelihood in refractory models

To begin, it is useful to discuss both the standard discretization approach and also the more refined method of (Citi et al., 2014). Both of these approaches begin by discretizing the observed continuous-time spike train process $\{t_i\}$ into a binary sequence ΔN_j , with a one in each bin (indexed by j) containing a spike time t_i , and a zero otherwise. The standard approach then applies a simple discretization of eq. (1) to obtain the approximation

$$l_{DR1}^\delta(t_1, \dots, t_{N(T)}) = \sum_{j=1}^{N_b} \Delta N_j \log(\lambda_j) - \lambda_j \delta. \quad (2)$$

Here λ_j denotes the CIF $\lambda(t|\mathcal{H}_t)$ evaluated at the center of the j -th bin, δ is the bin width, and $N_b = T/\delta$ is the number of bins. The approach of (Citi et al., 2014) instead uses the formula

$$l_{DR2}^\delta(t_1, \dots, t_{N(T)}) = \sum_{j=1}^{N_b} \Delta N_j \log(\lambda_j) - \left(1 - \frac{\Delta N_j}{2}\right) \lambda_j \delta. \quad (3)$$

Notice that the only difference between these two approaches is in the second term; as discussed in (Citi et al., 2014), the latter approximation is more accurate because it accounts for the loss of intensity due to refractoriness: roughly speaking, on average, spikes are in the center of bins, and if bins are small then the CIF over the second half should be zero. See (Citi et al., 2014) for full

¹Recently submitted independent work by K. Lepage makes a similar point (Lepage, 2014).

details; we use the abbreviation “DR,” for discrete Riemann, for each of these approaches, with DR1 corresponding to the standard method and DR2 for the approach of (Citi et al., 2014).

Our approach avoids the discretization of the observed spike train into a binary sequence ΔN_j , and instead works with the original continuous-time loglikelihood more directly. Define a τ -refractory point process as a process with an absolute refractory period of length τ , i.e., $\lambda(t|\mathcal{H}_t) = 0$ for $t \in (t_i, t_i + \tau]$, for each spike time t_i . In many applications, $\lambda(t|\mathcal{H}_t)$ is a smooth function of time t away from the spike times $\{t_i\}$. Thus it is sensible to break up eq. (1) into two terms: the first term, a sum over i we can compute directly, and the second term, an integral over $[0, T]$ that we must approximate numerically in most cases. Since there are discontinuities at each spike time t_i , we begin by breaking up this integral into $N(T) + 1$ terms, to avoid these discontinuities. Thus, defining $t_0 = 0$ and $t_{N(T)+1} = T$, for a τ -refractory point process we have

$$\int_0^T \lambda(t|\mathcal{H}_t)dt = \sum_{i=0}^{N(T)} \int_{t_i}^{t_{i+1}} \lambda(t|\mathcal{H}_t)dt = \int_0^{t_1} \lambda(t|\mathcal{H}_t)dt + \sum_{i=1}^{N(T)} \int_{t_i+\tau}^{t_{i+1}} \lambda(t|\mathcal{H}_t)dt. \quad (4)$$

Assuming $\lambda(t|\mathcal{H}_t)$ is smooth on each of these $N(T) + 1$ subintervals, we can now simply apply standard quadrature methods to each subinterval. (Of course, if we know a priori that $\lambda(t|\mathcal{H}_t)$ is discontinuous at additional time points, then we can break these integrals into more terms as necessary.)

We compare two such standard quadrature methods below: a simple trapezoidal rule with quadrature points evenly spaced between $t_i + \tau$ and t_{i+1} (abbreviated below as “CT,” for continuous trapezoidal) and Gauss-Lobatto quadrature (“GL,” with quadrature points defined below). Both of these are suitable for approximating the integral of a function f defined on an interval $[a, b]$ for which we know the endpoint values $f(a)$ and $f(b)$ (Dahlquist & Björk, 2008). In our case, by τ -refractoriness, we know that $\lambda(t_i + \tau|\mathcal{H}_{t_i+\tau}) = 0$ for each $i > 0$, and we have computed $\lambda(t_i|\mathcal{H}_{t_i})$ already in order to compute the discrete sum in eq. (1). Therefore, once we have computed $\lambda(t|\mathcal{H}_t)$ at $t = 0$ and $t = T$ we indeed have the values of our integrand at the endpoints of each of the intervals of interest. (Other quadrature rules are of course available, and may lead to better performance in some special cases; for example, we examined a Clenshaw-Curtis rule here as well (Trefethen, 2008), but found that the GL method performed better.)

Recall that a Gaussian quadrature is a numerical rule that approximates the integral of a function f defined on $[-1, 1]$ via the weighted sum $\sum_{j=1}^m w_{j,m} f(x_{j,m})$, with the quadrature points $x_{j,m}$ and (nonnegative) weights $w_{j,m}$ depending only on m , not on f . Thus, once the array of pairs $\{(x_{j,m}, w_{j,m})\}_{j=1\dots m}$ has been precomputed for all required values of m (this is a standard calculation²; see e.g. (Dahlquist & Björk, 2008) for further details), to apply the GL method to each of our integrands indexed by i , we simply need to apply a linear change of variables from $[t_i + \tau, t_{i+1}]$ to $[-1, 1]$, choose a value of $m = m_i$, and then evaluate the sum $\sum_{j=1}^{m_i} w_{j,m_i} f_i(x_{j,m_i})$, with f_i a suitably translated and rescaled version of $\lambda(t|\mathcal{H}_t)$.

For the purpose of fair comparison in the numerical analyses presented below, we allow each approximation method to evaluate the CIF a total of M times over the interval $[0, T]$, where

²In detail, the GL quadrature nodes $x_{j,m}$ are the roots of $(1 - x^2)P'_{m-1}(x)$, where $P_m(x)$ are the Legendre polynomials defined recursively by $(m + 1)P_{m+1}(x) = (2m + 1)P_m(x) - mP_{m-1}(x)$, $P_0(x) = 0$, $P_1(x) = 1$. Also, $w_{1,m} = w_{m,m} = 2/(m(m - 1))$ and for $1 < j < m$, $w_{j,m} = 2/(m(m - 1)P_{j-1}(x_{j,m})^2)$.

M is an accuracy parameter we vary systematically. (Since the quadrature points and weights $\{(x_{j,m}, w_{j,m})\}_{j=1\dots m}$ need only be computed once, the amortized cost of the GL method is exactly the same as that of the other three approximation methods considered here, if the same M is used for each method.) For the discrete methods (DR1 and DR2), $M = N_b$, the number of bins. For the continuous-time methods (CT and GL), we have found that an effective strategy is to allocate a minimum number M_{min} of quadrature points to each of our $N(T) + 1$ intervals ($M_{min} = 2$ or 3 works well in the cases we have examined, and for GL even $M_{min} = 1$ works in most cases, since this defaults to a triangular quadrature rule applied between $t_i + \tau$ and t_{i+1}), and then to allocate the remaining points so that m_i is roughly proportional to the length of the corresponding interval $t_{i+1} - t_i - \tau$.

Numerical results

For concreteness, we tested the performance of the four methods (DR1, DR2, CT, GL) using three different renewal processes, each with an absolute refractory period of length $\tau = 2$ ms. For the interspike interval (ISI) distributions of these three processes, we used a (τ -shifted) Rayleigh, inverse gaussian and log-normal. For each of these, 50 samples were drawn from the process, each with $T = 200$ s. Results are shown in Figure 1. In the top panels we show a sample of the CIF of each process, from $t = 0$ until the sixth spike time, $t = t_6$. In the center panel we show the approximated log-likelihoods for a single sample as a function of M , along with the corresponding exact values. Finally, in the bottom panel we show the median approximation error (\pm quartiles).

For the Rayleigh distribution (left column) both the CT and GL methods have negligible error: indeed, recall that GL is exact for polynomials of degree lower than roughly $2m_i$, where m_i is the number of quadrature points on the i -th interval, and in this case the CIF between spikes is linear. In all cases, we see clear differences in performance: the GL method is much more accurate (for sufficiently large M) than the continuous-time approach with trapezoidal quadrature (CT), which in turn is uniformly more accurate than the discrete time approximations DR1 and DR2. Indeed, the difference between the performance of the best continuous-time approach (GL) and the best discrete-time approach (DR2) is much larger than the difference between the two discrete-time approaches (DR1 and DR2), largely because the continuous-time approaches handle the discontinuities in the likelihood more accurately, and (as in the GL case) allow us to bring more sophisticated numerical integration methods to bear on the resulting smooth integrands, even in low firing rate regimes where refractory effects are less relevant. To summarize, the GL approach will be preferred over the standard discrete approaches whenever high accuracy approximations of the continuous time likelihood are required, and the CIF is piecewise smooth (and in particular when the CIF is well-approximated by a piecewise low-order polynomial).

3 Application to generalized linear models

We close by discussing the application of the methods discussed above to the generalized linear model (GLM) setting, where the CIF is specified in the form $\lambda(t|\mathcal{H}_t) = f(X_t\theta)$ for some known covariate vector X_t and an unknown parameter vector θ . Here $f(\cdot)$ is a nonnegative function

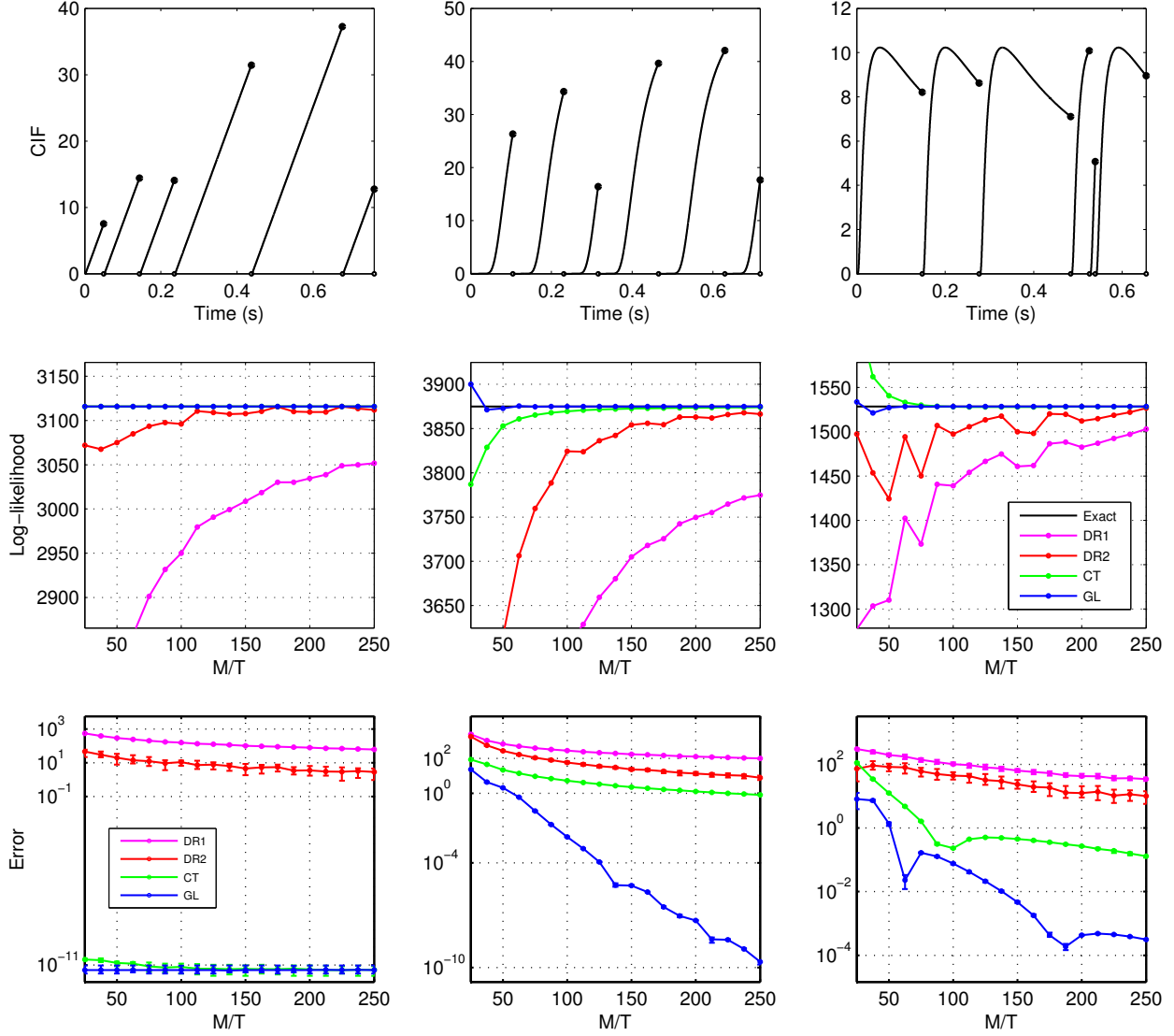


Figure 1: Comparing the four different likelihood approximation methods on three example renewal processes with absolute refractory periods of $\tau = 2$ ms. Left column: the ISI distribution was given by a Rayleigh distribution shifted by τ , with scale $\sigma = \frac{1}{10} \sqrt{\frac{2}{\pi}}$. Center column: Inverse gaussian, mean $\mu = \frac{1}{10}$ and shape $\lambda = 1$. Right column: Log-normal, shape $\sigma = 1$, and log-shape $\mu = -2.5$. Top row: CIF during the first six spikes of a single trial. Note that the CIF falls to zero after each spike. Parameters were chosen so that all the mean firing rates were between 10 and 15 Hz. Middle row: One sample of the approximated log-likelihood as a function of M/T , the total number of CIF evaluations per second, for each of the four methods (the two discrete time approximations, DR1 and DR2, the continuous time approximation based on Riemann sums, CT, and the Gauss-Lobatto quadrature method, GL). Bottom row: Median and quartiles of the absolute approximation error as a function of M/T (note log scale). The first plotted value of $M s^{-1}$ is roughly the smallest for which it is possible to distribute $M_{\min} = 3$ CIF evaluation points (not including $t_i + \tau$, where the CIF is known to be zero) into each integral. The GL approach clearly produces the best results.

which is typically assumed to be convex and log-concave, to ensure the concavity of the resulting loglikelihood with respect to the parameter θ (Paninski, 2004). If we want to enforce an absolute refractory effect of length τ , we can modify this model, e.g.:

$$\lambda(t|\mathcal{H}_t) = f(X_t\theta)r(t - t_{N(t)-})$$

where $t_{N(t)-}$ denotes the time of the last spike before t and $r(t)$ is a continuous, increasing function taking values in $[0, 1]$, with $r(u) = 0$ for $u \in [0, \tau]$.

An obvious question arises: can we utilize the proposed GL quadrature methods within standard likelihood optimization or MCMC approaches for inferring θ ? This turns out to be straightforward once we write the output of the GL approximation in the familiar form

$$\sum_{i=1}^M (N_j \log(\lambda(t_j|\mathcal{H}_{t_j})) - v_j \lambda(t_j|\mathcal{H}_{t_j}))$$

where $v_j \geq 0$ and t_j are the quadrature weights and nodes and $N_j = 1$ if t_j is a spike time and 0 otherwise. Really the only difference between this form of the loglikelihood and that used in the standard discrete-time approaches (whether DR1 or DR2) is that t_j and v_j are defined differently. In either case, this is a sum of concave functions of θ , hence a concave function, whose gradient and Hessian can be easily evaluated, and we can apply standard MCMC or iteratively reweighted least squares (IRWLS; Newton-type) methods to perform inference, using a number of available GLM packages. Similar points have been made earlier in the point process literature, e.g. Berman & Turner (1992), where related numerical approximations are introduced as more flexible alternatives to the standard approximations for the log likelihood based on binary time series.

To illustrate how better approximations provided by the GL quadrature can lead to improvements in inference quality, we examined a simple two parameter GLM with a piecewise linear refractory function $r(t)$. Results are shown in figure 2. The main conclusion is that GL quadrature approximate ML solutions are the fastest to converge to the limiting continuous-time ML values in this case, as a function of M .

The above framework relies on the fact that the quadrature points and weights $\{(x_{j,m}, w_{j,m})\}_{j=1\dots m}$ are independent of the integrand, and therefore independent of θ . More sophisticated adaptive quadrature methods are available that allocate quadrature points to regions in which the integrand is varying most quickly. In some cases (for example, in settings where the CIF is mostly constant as a function of time, interrupted by brief periods of high variance), these adaptive methods can lead to much more accurate approximations, given a fixed computational budget (defined by the parameter M). However, the resulting approximation depends on the integrand, and therefore in turn depends on θ , which typically leads to a discontinuous approximate loglikelihood as a function of θ (since the quadrature points may change discontinuously as a function of θ). This complicates the application of standard GLM inference methods. An alternative approach that could be useful in the context of maximum likelihood or maximum a posteriori approaches is to use a coarse-to-fine method: start with some crude estimate of θ , compute the corresponding CIF, compute the resulting quadrature points and weights adaptively using a small value of M based on this estimated CIF, and then perform a few steps of IRWLS, holding these quadrature points fixed.

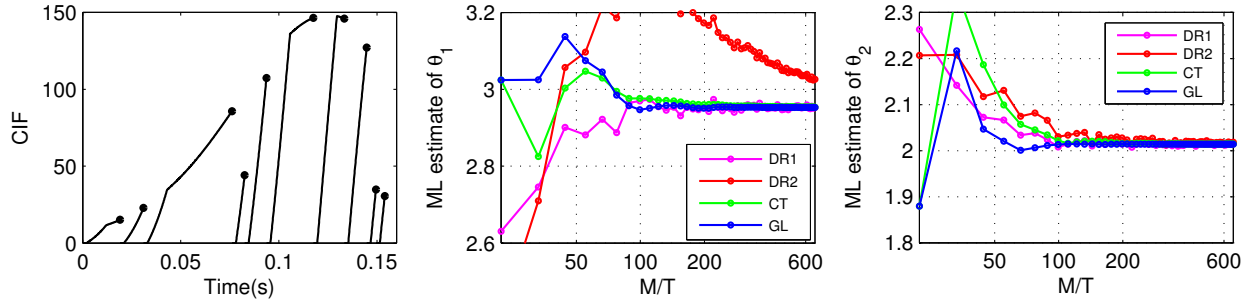


Figure 2: Comparing the approximate ML solutions provided by the different methods. A spike train was sampled for $T = 40s$ with CIF given by $\lambda(t|\mathcal{H}_t) = \exp(\theta_1 \sin(4\pi t) + \theta_2)r(t - t_{N(t)_-})$ with $r(t) = \max(0, \min(\frac{1}{\tau_2}(t - \tau_1), 1))$, $\tau_1 = 2ms$, $\tau_2 = 10ms$, $\theta = (3, 2)$. Left: CIF during the first ten spikes. Center and Right: Maximum likelihood approximate solutions of θ_1 and θ_2 using the different methods, for different values of M/T .

This provides a “coarse-scale” estimate of θ . Then this process can be repeated, increasing M (and therefore obtaining a refined estimate for θ) in an outer loop, holding the quadrature points fixed within each inner loop. We leave the exploration of this coarse-to-fine method for future work.

To help the reader get a better sense of the utility and implementation of the methods presented here, we have made available two Matlab examples. Essentially they reproduce the computations needed to make both figure 1 and figure 2. These can be downloaded at <http://stat.columbia.edu/~gonzalo/publications/>.³

Acknowledgements

We thank Josh Merel, Ari Pakman and the anonymous reviewers for helpful comments. Liam Paninski is supported by a NSF career award and by ARO MURI W911NF-12-1-0594.

References

- Berman, M. & Turner, T. R. (1992). Approximating point process likelihoods with GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1), 31 – 38.
- Brillinger, D. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 59(3), 189 – 200.
- Citi, L., Ba, D., Brown, E. & Barbieri, R. (2014). Likelihood methods for point processes with refractoriness. *Neural Computation*, 26(2), 237 – 263.

³ The code makes use of the Gaussian quadrature code of Greg von Winckel, available at <http://www.mathworks.com/matlabcentral/fileexchange/4775-legende-gauss-lobatto-nodes-and-weights/content/lglnodes.m>

- Dahlquist, G. & Björck, A. (2008). *Numerical methods in scientific computing, vol. 1*. Philadelphia:SIAM.
- Daley, D. J. & Vere-Jones, D. (2007). *An introduction to the theory of point processes: General theory and structure, vol. 2*. New York:Springer-Verlag.
- Lepage, K. (2014). Fast Maximum-Likelihood Estimation for Continuous-Time Neural Point Process Models. *Submitted*.
- Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4), 243 – 262.
- Trefethen, L. (2008). Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Rev.*, 50(1), 67 – 87.
- Truccolo, W. (2010). Stochastic models for multivariate neural point processes: collective dynamics and neural decodings. In Grün,S. & Rotter, S. (Eds.) *Analysis of parallel spike trains* (pp. 321–341). Berlin, Germany:Springer-Verlag.
- Truccolo, W. Eden, U., Fellows, M., Donoghue, J. & Brown, E. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2), 1074 – 1089.